



Analyse Factorielle Discriminante Multi-voie

Laurent Le Brusquet, Arthur Tenenhaus

► To cite this version:

Laurent Le Brusquet, Arthur Tenenhaus. Analyse Factorielle Discriminante Multi-voie. JdS 2013, May 2013, Toulouse, France. 6p. hal-00862329

HAL Id: hal-00862329

<https://hal-centralesupelec.archives-ouvertes.fr/hal-00862329>

Submitted on 16 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ANALYSE FACTORIELLE DISCRIMINANTE MULTI-VOIE

Laurent Le Brusquet & Arthur Tenenhaus

*E3S - Supélec Systems Sciences, 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France
prenom.nom@supelec.fr*

Résumé. L'analyse factorielle discriminante est étendue aux données multi-voie, c'est-à-dire aux données pour lesquelles plusieurs modalités ont été observées pour chaque variable. Les données multi-voie sont ainsi structurées en tenseur. L'extension proposée repose sur une modélisation des axes discriminants. Cette modélisation prend en compte la structure tensorielle des données. Les gains attendus par rapport aux méthodes consistant à construire un classifieur à partir de la matrice obtenue par dépliement du tenseur, sont une meilleure interprétabilité et un meilleur comportement vis-à-vis du sur-apprentissage, phénomène d'autant plus présent dans le contexte multi-voie que le nombre de modalités est grand. Un algorithme de directions alternées permet d'obtenir les axes discriminants. Les performances obtenues sur données simulées permettent de confirmer ces gains.

Mots-clés. Données multi-voie, analyse factorielle discriminante

Abstract. An extension of Fisher Discriminant Analysis to multi-way data is presented in this paper. In this multiway framework, individuals are characterized by variables which are measured at different conditions (e.g. scale, time,...) and yields dataset which is naturally structured in tensor. The proposed approach is based on a modelisation of the discriminant axis which take into account the tensor structure of the data. Expected improvements compared with unfolded methods are (i) better interpretability of the resulting model and (ii) better control of the over-fitting issue (even more striking issue when the number of conditions increase). An alternating directions algorithm is proposed for Multiway FDA and their performances are evaluated on simulated data.

Keywords. Multiway data set, Fischer discriminant analysis

1 Introduction

Ce papier propose d'étendre l'analyse factorielle discriminante proposée par Fisher (FDA) aux données multi-voie pour lesquelles les données sont un ensemble de K matrices de même taille (par opposition à l'analyse discriminante standard où les données sont représentées par une seule matrice). Plusieurs méthodes d'analyse de données, qu'elles soient spécifiques aux données multi-voie ou bien issues de méthodes d'analyse de données matricielles, ont été proposées afin de prendre en compte la structure tensorielle des données. L'article de Bro (2000) présente un riche aperçu de ces méthodes.

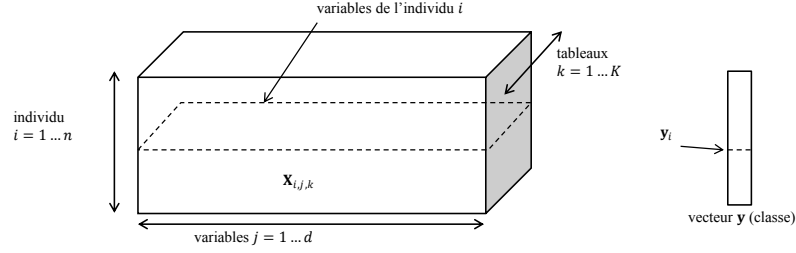


Figure 1: Données organisées selon un tenseur d'ordre 3 : chaque individu est représenté par K vecteurs de taille d .

Dans le contexte de la classification supervisée, plusieurs méthodes ont été proposées, souvent inspirées de méthodes classiques. Citons par exemple la méthode N-PLS (N-Partial Least Squares proposée par Bro (1996)), véritable extension de la méthode PLS pour gérer les tenseurs d'ordre ≥ 3 . Toujours avec un objectif de classification, ont été proposées des méthodes misant sur une réduction de la dimension des données qui tient compte de leur structure tensorielle. Une méthode de classification est ensuite appliquée aux composantes extraites. Parmi ces méthodes, citons la méthode PARAFAC proposée par Harshman (1970) ou la méthode STATIS proposée par L'Hermier des Plantes et Thiébaud (1977) qui sont des extensions de l'ACP aux données multi-voie.

Soient donc n résultats d'expériences, chaque expérience conduisant en l'observation de d variables observées selon K modalités différentes et d'une classe $c \in [1 \dots C]$ à prédire. Les données peuvent ainsi être rangées dans le tenseur $\{\mathbf{X}_{ijk}\}_{1 \leq i \leq n, 1 \leq j \leq d, 1 \leq k \leq K}$ d'ordre 3 représenté par la Figure 1, c'est-à-dire un ensemble de K tableaux de taille $n \times d$. Les classes observées sont, comme dans le cas matriciel, représentées par un vecteur \mathbf{y} de taille $n \times 1$. La multiplicité des tableaux conduit à un nombre total de variables de $K \times d$ donc potentiellement plus grand que le nombre n d'exemples. C'est particulièrement vrai lorsque l'on traite des données médicales pour lesquelles on a peu de patients sur lesquels on a pu observer de nombreuses variables (par exemple des images). Cet article s'inscrit pleinement dans ce contexte. Les enjeux sont à la fois la préservation de la structure tensorielle des données et la nécessité d'introduire de l'a priori structurant en raison du grand nombre de variables.

La méthode que nous proposons, nommée Multiway-FDA, est une extension de FDA présentée dans la section 2. La section 3 détaille le critère proposé pour répondre aux deux enjeux précédents ainsi que l'algorithme qui en découle. Les résultats obtenus sur des données simulées sont présentés en section 4.

2 Analyse factorielle discriminante

Une solution naïve consiste à déplier le tenseur en une matrice de taille $n \times (Kd)$ et de lui appliquer FDA. La matrice \mathbf{X}^u ainsi obtenue (^u pour unfolded) est composée de n vecteurs $\mathbf{x}_i = \text{vec}(\mathbf{X}_{i..})$ de longueur Kd .

L'analyse factorielle discriminante consiste à rechercher les $(C-1)$ projections de la forme $g(\mathbf{x}) = \mathbf{w}^\top \mathbf{x}$ où les $(C-1)$ vecteurs de poids \mathbf{w} sont des vecteurs de longueur Kd choisis tel qu'ils maximisent le rapport variance interclasse / variance intraclasse. Sans rentrer dans les détails de l'analyse discriminante (voir l'ouvrage de Hastie, Tibshirani et Friedman (2009)), notons seulement que ces variances sont des fonctions quadratiques de \mathbf{w} pouvant se mettre de la forme $\mathbf{w}^\top \mathbf{S}_{\text{Between}} \mathbf{w}$ et $\mathbf{w}^\top \mathbf{S}_{\text{Within}} \mathbf{w}$ avec : $\mathbf{S}_{\text{Between}} = (\mathbf{X}^u)^\top \mathbf{M}_{\text{Between}} \mathbf{X}^u$ et $\mathbf{S}_{\text{Within}} = (\mathbf{X}^u)^\top \mathbf{M}_{\text{Within}} \mathbf{X}^u$.

$\mathbf{M}_{\text{Between}}$ et $\mathbf{M}_{\text{Within}}$ sont des matrices $n \times n$ semi-définies positives ne dépendant que du vecteur \mathbf{y} . Les vecteurs \mathbf{w} s'obtiennent en résolvant le problème d'optimisation (1):

$$\mathbf{w} = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^\top \mathbf{S}_{\text{Between}} \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_{\text{Within}} \mathbf{w}} \quad (1)$$

qui conduit aux calculs des vecteurs propres de la matrice $(\mathbf{S}_{\text{Within}})^{-1} \mathbf{S}_{\text{Between}}$. Pour le contexte considéré ($n \ll K \times d$), une régularisation de la matrice est nécessaire et la matrice $\mathbf{S}_{\text{Within}}$ est remplacée par sa version régularisée $\mathbf{S}_{\text{Within}}^\mu = \mathbf{S}_{\text{Within}} + \mu \mathbf{I}$.

3 Multiway-FDA

La version multi-voie proposée s'inscrit pleinement dans le contexte $n \ll Kd$ puisqu'elle vise à la fois à imposer la structure des données au vecteur \mathbf{w} et à limiter les Kd degrés de liberté du vecteur \mathbf{w} . Elle consiste à choisir \mathbf{w} de la forme $\mathbf{w} = \mathbf{w}^K \otimes \mathbf{w}^d$ avec :

- \mathbf{w}^d : vecteur de poids pondérant l'importance des d variables,
- \mathbf{w}^K : vecteur de poids pondérant l'importance de chacun des K tableaux.

L'optimisation par rapport à \mathbf{w}^d et \mathbf{w}^K du critère (1) est menée à l'aide d'une procédure de type directions alternées. En effet, si l'on fixe l'un des deux vecteurs, les variances inter et intraclasse $\mathbf{w}^\top \mathbf{S}_{\text{Between}} \mathbf{w}$ et $\mathbf{w}^\top \mathbf{S}_{\text{Within}} \mathbf{w}$ s'interprètent comme les variances inter et intraclasse d'un critère FDA calculé sur des données matricielles (avec B/W = Between ou Within) :

$$\begin{aligned} \mathbf{w}^\top \mathbf{S}_{\text{B/W}} \mathbf{w} &= (\mathbf{w}^K \otimes \mathbf{w}^d)^\top (\mathbf{X}^u)^\top \mathbf{M}_{\text{B/W}} \mathbf{X}^u (\mathbf{w}^K \otimes \mathbf{w}^d) \\ &= (\mathbf{w}^d)^\top (\mathbf{X}^u (\mathbf{w}^K \otimes \mathbf{1}_d))^\top \mathbf{M}_{\text{B/W}} (\mathbf{X}^u (\mathbf{w}^K \otimes \mathbf{1}_d)) \mathbf{w}^d \end{aligned} \quad (2)$$

$$= (\mathbf{w}^K)^\top (\mathbf{X}^u (\mathbf{1}_K \otimes \mathbf{w}^d))^\top \mathbf{M}_{\text{B/W}} (\mathbf{X}^u (\mathbf{1}_K \otimes \mathbf{w}^d)) \mathbf{w}^K. \quad (3)$$

Algorithm 1 Calcul de l'axe principal d'une analyse Multiway-FDA :
 $(\mathbf{w}^{K(q)}, \mathbf{w}^{d(q)}) = \text{AxePrincipal}(\mathbf{X}^u)$

Require: $\epsilon > 0, \mathbf{w}^{K(0)}$

q \leftarrow 0

repeat

$$\mathbf{w}^{d(q+1)} \leftarrow \arg \max_{\mathbf{w}^d} \frac{(\mathbf{w}^d)^\top (\mathbf{X}^u (\mathbf{w}^{K(q)} \otimes \mathbb{1}_d))^\top \mathbf{M}_{\text{Between}} (\mathbf{X}^u (\mathbf{w}^{K(q)} \otimes \mathbb{1}_d)) \mathbf{w}^d}{(\mathbf{w}^d)^\top (\mathbf{X}^u (\mathbf{w}^{K(q)} \otimes \mathbb{1}_d))^\top \mathbf{M}_{\text{Within}} (\mathbf{X}^u (\mathbf{w}^{K(q)} \otimes \mathbb{1}_d)) \mathbf{w}^d}$$

$$\mathbf{w}^{K(q+1)} \leftarrow \arg \max_{\mathbf{w}^K} \frac{(\mathbf{w}^K)^\top (\mathbf{X}^u (\mathbb{1}_K \otimes \mathbf{w}^{d(q+1)}))^\top \mathbf{M}_{\text{Between}} (\mathbf{X}^u (\mathbb{1}_K \otimes \mathbf{w}^{d(q+1)})) \mathbf{w}^K}{(\mathbf{w}^K)^\top (\mathbf{X}^u (\mathbb{1}_K \otimes \mathbf{w}^{d(q+1)}))^\top \mathbf{M}_{\text{Within}} (\mathbf{X}^u (\mathbb{1}_K \otimes \mathbf{w}^{d(q+1)})) \mathbf{w}^K}$$

q \leftarrow q + 1

until $\|\mathbf{w}^{K(q)} - \mathbf{w}^{K(q-1)}\| < \epsilon$

return $(\mathbf{w}^{K(q)}, \mathbf{w}^{d(q)})$

Les matrices $\mathbf{X}^u (\mathbf{w}^K \otimes \mathbb{1}_d)$ et $\mathbf{X}^u (\mathbb{1}_K \otimes \mathbf{w}^d)$ sont des matrices de taille $n \times d$ égales à $\mathbf{X}^d = \sum_{k=1}^K (\mathbf{w}^K)_k \mathbf{X}_{..k}$ et $\mathbf{X}^K = \sum_{j=1}^L (\mathbf{w}^L)_j \mathbf{X}_{.j.}$. Optimiser le critère (1) selon \mathbf{w}^d (resp. \mathbf{w}^K) revient donc à considérer l'équation (2) (resp. (3)), i.e. réaliser une FDA sur la matrice \mathbf{X}^d (resp. \mathbf{X}^K). Ces FDA peuvent être réalisées à l'aide de version à noyau (voir le papier de Mika et al. (1996)) pour les applications où d et/ou K sont grands. L'algorithme (1) permet de calculer $(\mathbf{w}_1^K, \mathbf{w}_1^d)$ correspondant au maximum global du critère (1). Nous avons choisi d'identifier les $C - 2$ axes suivants (représentés par les vecteurs $(\mathbf{w}_s^K, \mathbf{w}_s^d)$, $s = 2 \dots C - 1$) en optimisant le critère (1) sous la contrainte que chaque vecteur \mathbf{w}_{s+1}^K soit orthogonal aux vecteurs précédents. Ce choix présente l'avantage de pouvoir utiliser à nouveau l'algorithme (1) précédent. En effet, supposons calculés les s premiers axes et soit \mathbf{P} une matrice de projection de \mathbb{R}^K vers le sous-espace de dimension $K - s$ orthogonal aux $\mathbf{w}_{s'}^K, s' = 1 \dots s$. En réécrivant les équations (2) et (3) en fonction du vecteur $\mathbf{w}' = \mathbf{P} \mathbf{w}_{s+1}^K$ de longueur $K - s$, il est possible de montrer qu'appliquer l'algorithme (1) au tenseur \mathbf{X}' , de dimension $(n, d, K - s)$ défini par $\mathbf{X}'_{ij.} = \mathbf{P}^\top \mathbf{X}_{ij.}$, $i = 1 \dots n$, $j = 1 \dots d$, permet de calculer $(\mathbf{w}', \mathbf{w}_{s+1}^d)$ puis $\mathbf{w}_{s+1}^K = \mathbf{P}^\top \mathbf{w}'$. En appliquant axe par axe cette démarche, l'algorithme (2) permet de calculer les $(C - 1)$ axes.

4 Exemple illustratif

La méthode a été appliquée à un jeu de données simulés avec $(n, d, K, C) = (80, 100, 20, 3)$. Pour un individu et un tableau données, les d variables $X_{i.k}$ sont générées à l'aide d'un polynôme évoluant lentement en fonction du n° du tableau (Figure 2). Une FDA sur le tenseur déplié ainsi que Multiway-FDA ont été appliqués.

À l'instar de FDA, Multiway-FDA permet de représenter les individus sur le plan dis-

Algorithm 2 Multiway-FDA, calcul de toutes les composantes

```

 $\mathbf{X}' \leftarrow \mathbf{X}, \mathbf{P} \leftarrow \mathbb{1}_K$ 
for  $s = 1$  to  $C - 1$  do
   $(\mathbf{w}'^K_s, \mathbf{w}^d_s) \leftarrow \text{AxePrincipal}(\mathbf{X}')$ 
   $\mathbf{w}^K_s \leftarrow \mathbf{P}^\top \mathbf{w}'^K_s$ 
   $\mathbf{P} \leftarrow$  matrice de projection de  $\mathbb{R}^K$  vers le sous-espace orthogonal aux  $\mathbf{w}^K_{s'}, s' = 1 \dots s$ 
   $\forall i \in [1, n], \forall j \in [1, d], \quad \mathbf{X}'_{ij.} = \mathbf{P} \mathbf{X}_{ij.}$ 
end for
return  $(\mathbf{w}^K_s, \mathbf{w}^d_s), s = 1 \dots C - 1$ 

```

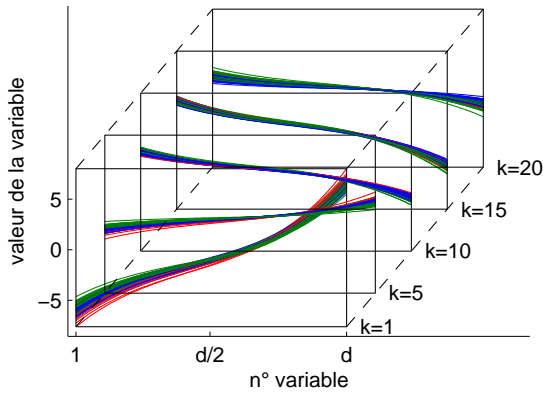


Figure 2: Exemple avec $K = 20$ tableaux et $C = 3$ classes. Chaque classe est représentée par une couleur différente. Seuls 5 tableaux sont représentés.

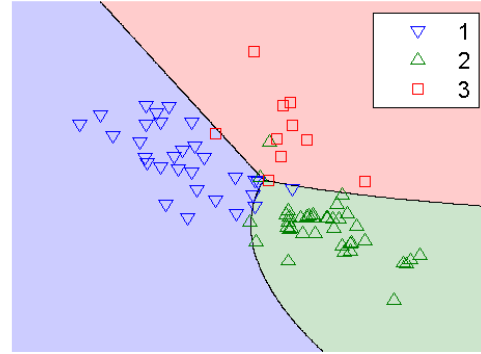


Figure 3: Représentation des n individus dans le plan principal de Multiway-FDA $(\mathbf{t}_1, \mathbf{t}_2)$.

criminant $(\mathbf{t}_1 = \mathbf{X}^u(\mathbf{w}_1^K \otimes \mathbf{w}_1^d), \mathbf{t}_2 = \mathbf{X}^u(\mathbf{w}_2^K \otimes \mathbf{w}_2^d))$. La Figure 3 montre les individus projetés ainsi que le classifieur obtenu par une analyse discriminante quadratique de \mathbf{y} sur $\mathbf{t}_1, \mathbf{t}_2$. Les Figures 4 et 5 permettent de comparer les modèles obtenus par FDA et Multiway-FDA : Multiway-FDA permet de pondérer séparément l'influence des variables et des tableaux sans perdre en capacité de prédiction. Une validation croisée par Leave-One-Out a en effet été menée et les taux de classification sont comparables pour les deux approches (86,3 % pour Multiway-FDA et 87,5 % pour FDA).

5 Conclusion

L'extension de l'analyse factorielle discriminante aux données multivoie est proposée. Elle prend en compte la structure tensorielle des données. Les gains attendus sont une meilleure interprétabilité des axes discriminants obtenus et un possible meilleur taux de classification (la prise en compte de la structure des données permettant de contrer le

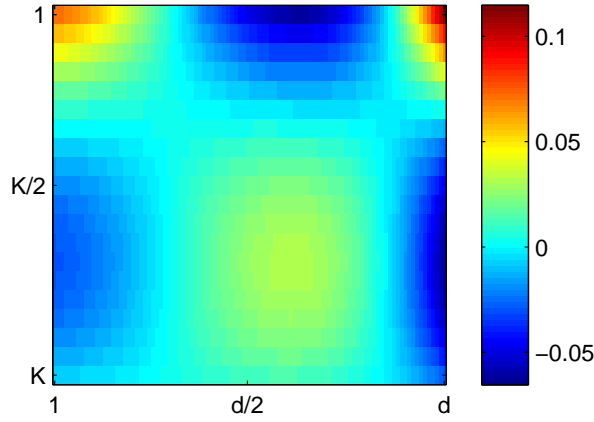


Figure 4: Premier vecteur de poids de taille $K \times d$ pour la FDA calculée avec \mathbf{X}^u .

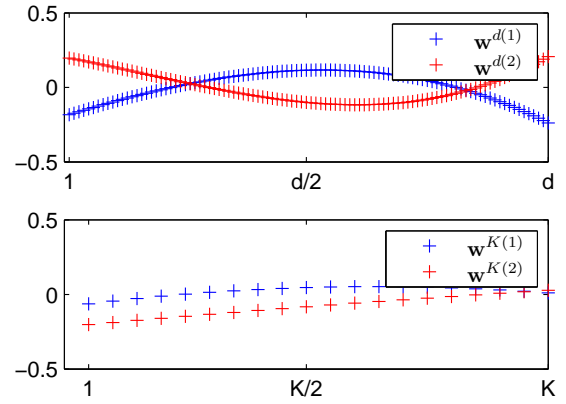


Figure 5: Vecteurs \mathbf{w}_s^d et \mathbf{w}_s^K calculés par Multiway-FDA pour $s = 1$ et 2.

phénomène de sur-apprentissage dans le contexte d'un faible nombre d'individus observés). Par souci de simplicité d'écriture, cette extension a été proposée dans le cadre de tenseurs d'ordre 3. Elle se généralise cependant à des tenseurs d'ordre quelconque. Dans un contexte de variables en nombre toujours plus grand, et afin de privilégier l'interprétabilité des axes discriminants, il est envisagé de construire des versions régularisées de la matrice $\mathbf{S}_{\text{Within}}$ de la forme $\mathbf{S}_{\text{Within}}^\mu = \mathbf{S}_{\text{Within}} + \mu \mathbf{R}$, \mathbf{R} ayant elle aussi une structure tensorielle.

Bibliographie

- [1] Bro, R. (2000), Multi-way Analysis in the Food Industry - Models, Algorithms, and Applications *ICSLP Proceedings*.
- [2] Bro, R. (1996), Multiway calibration. Multilinear PLS, *Journal of Chemometrics*, vol. 10, n. 1, 47–61.
- [3] Harshman, R. A. (1970), Foundations of the PARAFAC procedure: Model and conditions for an explanatory multi-mode factor analysis, *UCLA Working Papers in phonetics*, 16.
- [4] L'Hermier des Plantes, H. et Thiébaud, B. (1977), Etude de la pluviométrie au moyen de la méthode STATIS, *Revue de Statistique Appliquée*, 25, 57–87.
- [5] Hastie, T., Tibshirani, R. and Friedman, J. (2009), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer.
- [6] Mika, S., Rätsch, G., Weston, J., Schölkopf, B. et Müller K.-R. (1999), Fisher Discriminant Analysis with Kernels.